



Building Repositories: Why DSpace?

*Ann Devenish, WHOAS Project Manager
Digital Commonwealth Conference
College of the Holy Cross
25 October 2007*



Agenda

Determining factors in selecting the IR platform

Where we are at present

Lessons learned along the way



MBLWHOI Library

Joint library serving the Woods Hole science institutions:

Marine Biological Laboratory

National Marine Fisheries Service/Northeast Fisheries Science Center 🇺🇸

SEA Education Association

U.S. Geological Survey Science Center for Coastal and Marine Geology 🇺🇸

Woods Hole Oceanographic Institution

Woods Hole Research Center



“I have a dream”*

we're going to build an e-print server

Cathy Norton, director MBLWHOI Library

... a what?

Ann Devenish
September 2002

(*with apologies to Martin Luther King, Jr.)



Dream shaped by desire

- *Use a widely accepted metadata scheme*
- *Content has to be easily discovered and easily accessed*
- *Content has to be open access*
- *Provide for long-term preservation and/or content migration as formats “die”*



Dream shaped by intent

- *Host all types of digital formats, including large data sets*
- *Reflect separate institutions with potentially differing needs, mandates, content, audiences*
- *Identified with the Library*



Dream shaped by reality

- *No new dollars; no new staff*
 - *“Herbarium” model*
creative adaptation of available in-house software, or
 - *Free software*
- *KISS (keep it simple)*
 - *No software development*
 - *Minimum customization by IT staff*
- *No institutional directives for self-archiving*
 - *Voluntary participation by authors*



DSpace features

- *Free and open source platform from a trusted source (MIT & HP)*
- *Supports qualified Dublin Core metadata*
- *Handle System (from CNRI) to assign and resolve persistent identifiers ~ <http://hdl.handle.net/1912/> ...*
- *Supports OAI PMH (open archives initiative protocol for metadata harvest) as a data provider*



DSpace features, cont.

- *Wide variety of user roles: Anonymous visitor → Full administrative capabilities*
- *“Communities” have their own identity and content guidelines reflecting differing policies and practices*
- *Completely customizable: run as is, or modify and extend**

** Java, Apache Ant, PostgreSQL or Oracle, Jakarta Tomcat*



DSpace “plus”

Large and growing community of users provides advice and assistance, and shares development

Recent enhancements:

- *Item importer for large files and/or batches of files*
- *Full text extraction and searching for MS Word, PDF, HTML, text documents*
- *Statistics*
- *LDAP authentication*
- *Configurable displays*
- *RSS feeds*
- *Support for Creative Commons licenses*



“If you build it, he will come” Recruiting content

Push (“unsolicited” contributions) from researchers:

<3% of total Woods Hole content in WHOAS

*<4% of the total number of articles
estimated to be eligible have been added
to PMC*

Report on the NIH Public Access Policy, Jan. 2006



Articles published 2005 – date

Pull (harvest) published version:

If a permitted use by the copyright holder (usually the publisher)

Pull (solicit) pre-prints/post-prints:

Identify ~ 1720+ articles

Verify ~ 52% may be eligible for hosting in an IR as pre-print or post-print

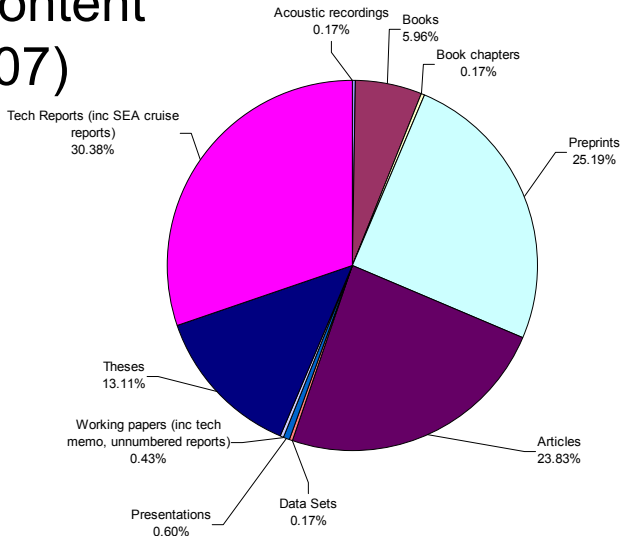
Solicit ~ 32% response rate

Result: Loaded ~ 435+ pre-prints & published articles



Current content (1880-2007)

≈1175 records
(Oct. 4, 2007)





Lessons learned

Involve library and IT colleagues, early and often

Be flexible

Be patient

Be persistent

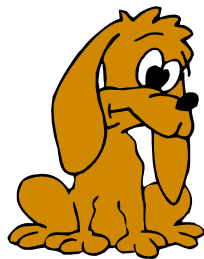
Expect the unexpected

Be prepared to un-do and re-do

Know your limitations

Mistakes are okay

Look forward



“There's no such thing
as a ‘free’ puppy”



Learn more about DSpace

- *DSpace.org*:
 - <http://www.dspace.org/>
- *DSpace wiki*:
 - http://wiki.dspace.org/index.php/Main_Page



Would we make the same choice again?

You might very well think that; I couldn't possibly comment ...

Francis Urquhart
(Michael Dobbs)



Other questions?

Ann Devenish
Publishing Services
WHOAS Project Manager
MBLWHOI Library
Woods Hole, MA 02543
adevenish@who.edu
508.289.2865